Dear Sir:

The SIGAPL Index of Technical Articles

ACM SIGAPL intends to publish an annual index of technical articles about APL; our plan being to gather material during 1989 for the first issue, which will cover the calendar year of 1989.

I am writing to you to announce this project and to enlist the help of your group in preparing the index; we want our coverage to be as broad as possible and want to encourage contributions from as many sources as we can - if you know about something then pass the information on - it is far better that we hear about the same item several times than not at all. Everyone who assists in this project will be acknowledged.

The SIGAPL Executive Board sees this project bringing many benefits, which will accrue for the APL community as a whole, among them:

> Better awareness of the breadth and depth of APL material available, both for the APL community and everyone else.
>
> Better-researched papers, eliminating duplicated effort and promoting development of ideas.
>
> Increased circulations for APL journals, as people realise the value of the material they contain.

The first step is obviously to ensure that we know about the contents of journals published by all the specialist APL groups (and vendors), we will expand coverage as we become aware of more sources of information - the only restriction on this exercise is that this will be an index of technical articles and not general articles.

As I envisage it at the moment we will need the following about each article:

> Title
> Author
> Citation (journal title, volume and issue, pages)
> Language of text
> Subscription/backorder address for journal
> keywords

In the interests of uniformity the ideal situation would be for me to receive issues of your journal which contain relevant articles, however should this not be possible I am happy to receive the above items. We will be reporting on the progress of this project through the pages of APL Quote Quad; if there's anything which you think would increase the effectiveness of the index (or reduce the effort needed to compile it) then please let me know.

I look forward to hearing from you.

Dick Bowman
ACM SIGAPL
c/o CEGB
85 Park Street
London SE1
England

Algorithm

Algorithm 179

The Jurkevich Periodogram

Glenn Schneider

The analytical and computational methods available for determining periodicities in observational data are many and varied. Most of these techniques such as the discrete or fast Fourier transformation [1], or harmonic analysis [2], require that the data are regularly spaced and uninterrupted. That is to say, the sampling interval must be constant and no data samples can be missed. These constraints cannot always be met in the acquisition of real data. One rather simple approach to period determination, for which these constraints do not apply, is the period search method of Jurkevich [3].

The Jurkevich search technique tests a range of periods and ascribes a level of confidence (the "scatter statistic", [4]) for each period tested. For each trial period, *P*, the search procedure collapses the data into a single cycle based on the phase of each data point. For a vector of observations taken at monotonically increasing times, *T*, the phase of each point is simply `1|(T-T[1])÷P`. A graph of the collapsed data is referred to as a phase diagram. If the data are truly periodic then the points on the phase diagram will fall neatly onto a cyclical curve, for the correct period. For an incorrect period, the phase diagram will resemble a scatter plot. The Jurkevich search procedure seeks to minimize the scatter about a cyclical curve in the phase diagram. As the scatter is reduced, the likelihood that the period under test is the correct one increases.

In order to quantitatively define the degree of scatter, the Jurkevich method divides the phase diagram into a number of bins, *J*. The means of the data points which fall into each of the bins, *M*, are computed. For incorrect trial periods, the data points will be randomly distributed in each of the bins, so that `M+.*2` approaches zero. Conversely, as the true period is approached there will be a systematic displacement of the points in some (or all) of the bins, above or below the mean value. Hence, `M+.*2` approaches a maximum as the true period is approached. The scatter statistic, *S*, is defined as `N+.×M*2`, where *N* is a vector of length *J*, containing the number of points in each bin.

The APL function, *JURK*, listed below, will perform a Jurkevich search on an observational data set. The left argument, *P*, is a four element numeric vector. `P[1 2]` specifies the starting and ending periods to be searched, respectively. `P[3]` specifies the number of Jurkevich bins (*J*) into which the phase-reordered data are to be distributed into. (For most cases a value of 3 is sufficient). `P[4]` is the increment in the periods to be searched, within the range specified by `P[1 2]`. The right argument, *D*, is a two column matrix containing the observational data. The first column gives the times of the observations, and the second column the corresponding data values. The first column must contain only integral values. Therefore, if the data were originally timetaged in fractional units, then appropriate scaling must be applied to `D[;1]`. For example, for data recorded with timetag values in fractional seconds, the times should be expressed as integral milliseconds, or microseconds, depending upon the precision required.

```
      ∇ S←P JURK D;L;M;N;⎕IO
[01]    ⍝P=START, END, NUMBER OF BINS,
                                  INCREMENT
[02]     ⎕IO←1
[03]     S← 0 2 ⍴0
[04]    J1:L←(⍳P[3])∘.=1+⌊(P[1]|D[;1])
                              ÷⌊+/P[1 3]
[05]     N←⎕CT⌈+/L
[06]     M←(L+.×D[;2])÷N
[07]     S←S,[1] P[1],N+.×M*2
[08]     →(≤/P[1 2]←P[1 2]+2↑P[4])/J1
[09]     S←S,(S[;2]÷⌈/S[;2])*2
      ∇
```

Listing of the APL function *JURK*.

JURK returns a three column matrix tabulating the periods searched (column 1), and the scatter statistics and indices coresponding to each of the trial periods (columns 2 and 3). The scatter indices are defined as the square of the scatter statistics normalized to the largest scatter statistic computed for all periods. The period for which the scatter index is equal to 1 is most likely the true period (assuming that the true period fell within the range of the periods searched).

As an illustrative example consider the synthetic data set listed in Table 1, and shown graphically in Figure 1. This data set contains only a small number of points (42), to demonstrate that the Jurkevich method can recover periodicities even in rather limited data sets. In practice, one normally would have a larger data volume. This particular data set simulates readouts from a data acquisition inststrument with the following characteristics; two successive readouts occur approximately every nineteen (± 0.5) minutes, spaced approximately two minutes (± 10 seconds) apart. The asynchronous nature of the data acquisition timing scheme simulates a possible mode of operation of an actual data acquisition instrument (the International Ultraviolet Explorer satellite echelle spectrograph). The acquisition times were generated stochastically based on the above timing rules. The synthetic data were generated from a sine function with a period of 157 seconds. The resulting synthetic data have a mean sampling interval of 1149.3 (± 27.13) seconds, with paired samples spaced an average of 123.38 (± 9.80) seconds apart. Gaussian noise was added to these data, with a distribution function whose 1 sigma noise level was 10 percent of the corresponding data values.

The function *JURK* was employed in order to extract the "unknown" period of the synthetic data. Trial periods in the range of 3 to 400 seconds were tested by executing:

```
3 400 3 1 JURK DATA,
```

where *DATA* was the matrix listed in Table 1. In addition, to serve as an example of the numerical output produced by *JURK*,

```
145 165 3 1 JURK DATA
```

was executed and the results tabulated in Table 2.

The Jurkevich periodogram shown in Figure 2 is a plot of the scatter indices for trial periods of 3 to 400 seconds using the test data set. The maximum scatter index found is seen at 156.94 seconds, which is nearly identical to the period of the synthetic data. A plot of these data, collapsed into a phase diagram with a period of 156.94 seconds is shown in Figure 3a. The scatter in the curve is only a result of the Gaussian noise added to the synthetic data. Visual inspection of this phase diagram indeed indicates that true period was found.

A number of other non-insignificant peaks, however, can be seen in the periodogram. These "false peaks" arise from two different effects; aliasing and harmonics. If the data subjected to processing by *JURK* were sampled in a periodic or quasi-periodic manner, or if repetitive data dropouts occurred during sampling, then a spectrum of false peaks will arise due to aliasing. Periodic sampling dropouts may be imposed on the observations from a variety of sources. For example, ground based astronomical observations are subject to daily interruptions, when the object under study drops below the horizon, or data acquisition is interfered with due to twilight or sunrise.

Aliasing is essentially a beating effect between the true period and the sampling period. Aliased peaks may be expected at periods given by `÷(÷P)-N÷I`, where *P* is the true period (as determined by a maximum scatter index), *N* is a vector of integral alias numbers (analogous to harmonic numbers) and *I* is the sampling interval. The synthetic data presented here were sampled in the quasi-periodic manner previously discussed. The spectrum of false peaks, aliased from the most likely period of 156.94 seconds and the 1149.3 second mean sampling interval for integral alias numbers -8, -3, -2, -1, +1, +2, +3 are seen at 75.01, 111.29, 123.13, 138.08, 181.71, 215.97 and 266.79 seconds. False peaks arising from the aliasing of the pairwise mean sampling interval of 123.38 seconds can be seen at 21.34 and 32.59 seconds.

Harmonics (multiples and submultiples) of both the true period and the principle alias spectrum will result in additional false peaks. For example, the first subharmonics of the true peak and of the 181.70 false peak can be seen at 313.95 and 363.41 seconds in Figure 2.

Phase diagrams of the synthetic data, with the periods indicated by the two largest false peaks (138.08 and 181.71 seconds), are shown in Figrues 3b and 3c. It is readily apparent, from the scatter in these plots (compare this to 3a), that these trial periods clearly are incorrect.

In the above discussion, the periods indicated by the false peaks are quoted to a precision greater than possible by performing a search with an increment of one second. Indeed, once the periodicities of interest (large scatter index) were found, *JURK* was rerun with small search windows (2 seconds) and an increment of 0.01 seconds centered on these periods. While this would normally not be done for the false peaks, it points out that an initial search should cover a wide range of periods with a course search step. Once the likely true period is found a much finer search should be performed about that period to refine its value.

Data sets which possess multiple periodicities may be processed repetitively by the Jurkevich search procedure in order to determine the component periods, in the following manner. After the principle period is found by an initial search, the data may be fit by a conventional method, such as least squares, to a periodic curve (i.e. a sinusoid) of the indicated period. The vector of residuals, arising from subtracting the fit points from the observational data, can then be searched by the Jurkevich procedure for a secondary period. This process, known as "prewhitening", can be iterated to extract multiple periodicities. Each iteration will produce its own spectrum of false peaks in the periodogram. The relationship between the Jurkevich periodogram peaks and the Fourier transformation coefficients is discussed in detail by Swingler [5].

A knowledge of the periodicities contained in observational data often is required before further analysis can be carried out. As a working tool, the Jurkevich period search method is extremely useful for finding approximate periods. The error in the determined periods depends upon the number of data samples, sampling intervals, and noise in the data, but usually is much less than one percent of the true value.

References

[1] Mayer, A., *The Fast Fourier Transform Implemented in APL*. APL Quote Quad, Vol. 16, No. 2 (1986), 12-16.

[2] Schneider, G., *Fourier Transformation by Harmonic Analysis*. APL Quote Quad, Vol. 11, No. 2 (1982), 19-21.

[3] Jurkevich, I., *A Method of Computing Periods of Cyclic Phenomena*. Astrophysics and Space Science, Vol. 13, No. 1 (1971), 154-167.

[4] DuPuy, D.L. and Hoffman, G.A., *A Jurkevich Period Search Program*. International Amateur-Professional Photoelectric Photometry Communication, No. 20 (1985), 1-17.

[5] Swingler, D.N., *A Relationship Between the Jurkevich Periodogram and the Fourier Transform Spectral Estimator*. Astronomical Journal, Vol. 90, No. 4 (1985), 675-679.

Astronomy Programs
Computer Sciences Corporation
Space Telescopy Science Institute
John Hopkins University
Homewood Campus
Baltimore, MD 21218
USA

TABLE 1.
SAMPLE INPUT DATA: 42 OBSERVATIONS

TIME	VALUE	TIME	VALUE
1176	0.0600	12824	-0.9089
1297	0.9975	13848	0.9582
2352	-0.1198	13975	0.0800
2484	-0.9003	14968	0.8524
3466	0.4620	15085	0.4971
3579	-0.9582	16148	-0.7959
4638	-0.2572	16276	-0.8726
4741	0.9460	17260	-0.3896
5754	-0.8078	17395	-0.9682
5886	0.0600	18403	0.9780
6911	0.1198	18535	0.3524
7030	-0.9856	19595	-0.9323
8078	0.2957	19726	-0.7836
8200	0.9916	20734	0.3896
9215	-0.9393	20848	-0.9689
9327	0.5483	21864	0.9975
10370	0.3147	21991	0.4261
10479	-0.9995	23029	-0.9089
11553	-0.5144	23166	-0.3336
11682	0.5483	24165	-0.4971
12688	-0.9170	24280	-0.8078

TABLE 2.
OUTPUT: SCATTER STATISTICS AND INDICES

p	$S(p)$	$I(p)$
145	3.2323	0.0277
146	1.6041	0.0068
147	1.8210	0.0088
148	1.4523	0.0056
149	1.5155	0.0061
150	2.7135	0.0195
151	2.0489	0.0111
152	1.1180	0.0033
153	1.1237	0.0033
154	2.7615	0.0202
155	1.1557	0.0035
156	1.1070	0.0032
157	19.4274	1.0000
158	1.1143	0.0033
159	2.2761	0.0137
160	2.9272	0.0227
161	2.6287	0.0183
162	0.7964	0.0017
163	0.8760	0.0020
164	1.0078	0.0027
165	1.1218	0.0033

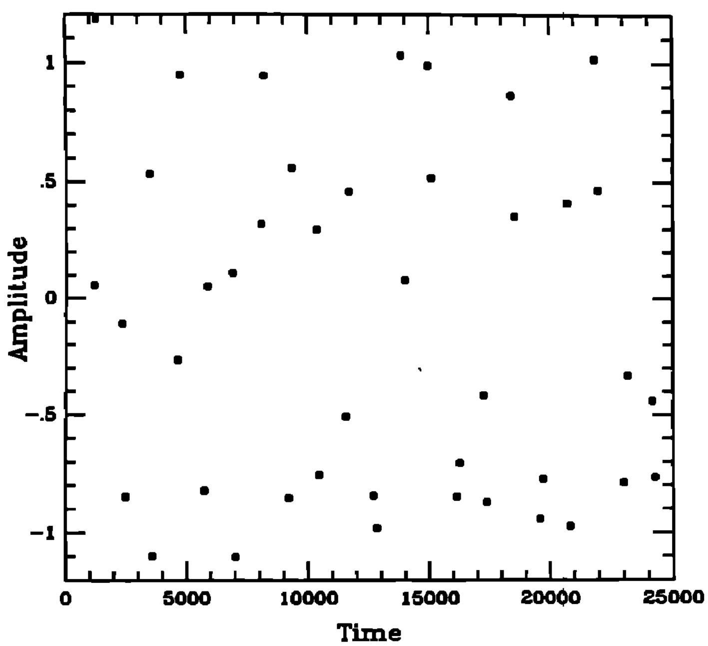

Figure 1.
Plot of the synthetic data set processed by *JURK* as discussed in the text

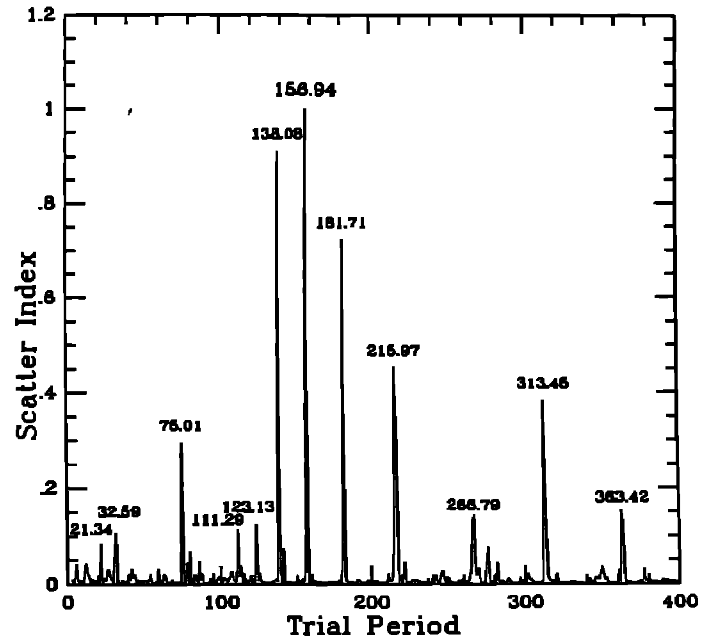

Figure 2.
The periodogram of the scatter indices determined by *JURK* for the synthetic data set for trial periods in the range of 3 to 400 seconds

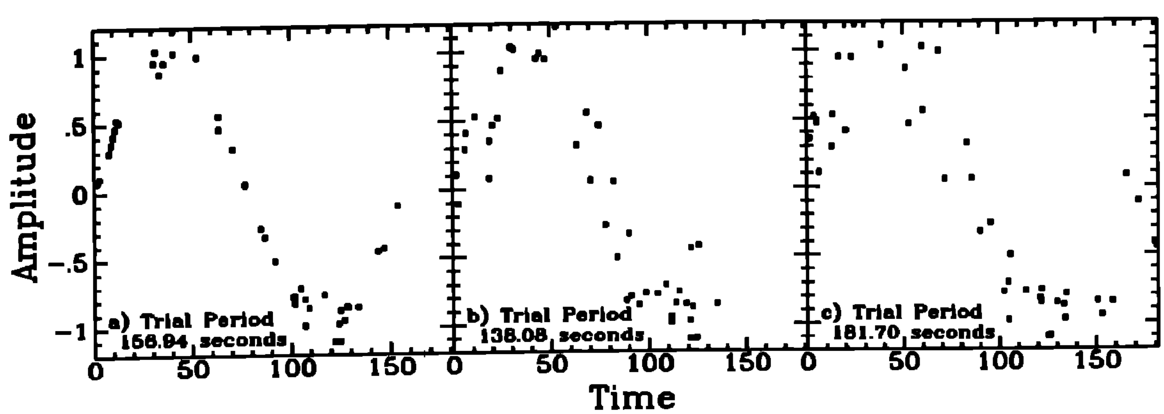

Figure 3.
Phase diagrams of the synthetic data set for trial periods of a) 156.94 seconds, b) 138.09 seconds, c) 181.71 seconds